**Redes Neurais Artificiais, Deep Learning e Algoritmos Genéticos**

# Checkpoints 02 e 03

|  |
| --- |
| Resolvam o problema de dados abaixo **em grupos de 4 pessoas** utilizando redes neurais artificiais para modelar o problema. Submetam um *jupyter notebook* no Teams.  **Data limite**: 26/maio/2022  **Importante**: coloquem o nome completo e RA de todos os membros do grupo em alguma célula do notebook. Apenas um representante do grupo deve submeter o trabalho. |

## Contexto

Vamos utilizar uma base de dados de imóveis de São Paulo para construir modelos preditivos baseados em redes neurais artificiais para *prever o preço de venda* de tais imóveis.

## Dataset

Nosso dataset está disponível no **Kaggle**: <https://www.kaggle.com/argonalyst/sao-paulo-real-estate-sale-rent-april-2019>

O arquivo é um CSV, disponível neste endereço: <https://tinyurl.com/alugueis-sp-df> , o qual pode ser lido diretamente pelo Pandas!

import pandas as pd

imoveis = pd.read\_csv('https://tinyurl.com/alugueis-sp-df')

imoveis.head(5)

Esta é uma amostra dos dados. O entendimento das features é praticamente direto pelos seus nomes, mas recomendo que consultem a página do Kaggle para maior detalhamento de cada campo. Lembrem-se: o entendimento dos dados faz parte do processo de data science!



## Desafio

Filtrem os dados para usar apenas informações dos aptos que estão à venda (*sale*). A partir desta filtragem, proponham uma rede neural que infere o valor de venda do imóvel baseado em suas características.

**O que é para ser feito?**

* Construam 4 modelos DISTINTOS variando dados e/ou arquiteturas. Comparem as métricas de avaliação e analisem os resultados;
* Para cada modelo, usem quantas e quais features acharem melhor;
* Tragam uma validação externa a partir de 5 novos apartamentos, cujos dados podem ser consultados em sites de imobiliárias (ex: Loft, QuintoAndar, ZAP Móveis, Viva Real, etc). Ou seja: acessem os sites das imobiliárias e escolham 5 aptos quaisquer para venda, coletando as mesmas infos da base e avaliem o comportamento dos modelos nestes aptos. O que você observou? Comentem os resultados.

**O que deve ser entregue?**

* Um jupyter notebook para o grupo, contendo:
  + o nome e RA dos integrantes e
  + todo o processo de análise de dados, desde a leitura da base, passando pela exploração, entendimento, limpeza, etc até chegar na modelagem final e sua avaliação (métricas e avaliação externa)
* Usem também [células markdown](https://blog.da2k.com.br/2015/02/08/aprenda-markdown/) para documentar o processo e incluir comentários pertinentes sobre suas decisões e principais análises conclusões.

**Como o trabalho será avaliado?**

* Lembre-se dos passos do pipeline de machine learning para uso do sklearn! Precisamos ler a base, limpá-la, pensar em quais dados serão incorporados nos modelos, padronizar dados, etc.
* A correção do checkpoint será feita mediante os seguintes critérios:
  + 1. emprego correto dos conceitos vistos em sala de aula (uso do dataset, aplicação do sklearn, métricas pertinentes e suas interpretações, etc)
    2. completute de atendimento ao pipeline de machine learning
    3. qualidade da compreensão e exploração da base
    4. qualidade/profundidade da limpeza dos dados
    5. existência de 4 modelos distintos sendo averiguados
    6. escolha de dados pertinentes para uso nos modelos
    7. formas de padronizar os dados selecionados
    8. construção da base de validação externa
    9. análise/justificativas dos resultados (validação interna e externa)
    10. apresentação dos resultados

**Qual é o escopo de cada trabalho?**

* CK #2 : o notebook com o trabalho desenvolvido
* CK #3 : apresentação do grupo, em sala de aula, no dia 26/maio/2022

Bom trabalho! Divirtam-se!